Create and Populate Tables in Hive

|  |  |
| --- | --- |
| **MySQL Database** | lanier |
| **MySQL Table** | device |
| **HDFS Path** | lanier/webpage |
| **Data File** | ~/materials/data |

**In these exercises you will define Hive tables to model and view data**

**in HDFS.**

Preparing Test Data

We need to prepare HDFS for the lesson. To do so, please follow the instructions below. Will will start by added the appropriate data from our data file, into HDFS using the “put” command.

1. cd ~/materials/data
2. hdfs dfs -put webpage /lanier

**Change Hive Metastore Parameter in Ambari**

This step is equally as important. Before continuing on, please make the following changes to your Hive Metastore, through Ambari by following along with the instructions below.

1. **Click Hive** > **Configs** > **Advanced**
2. Scroll down till you see “custom hive-site” and expand the drop down menu by clicking on triangle.
3. Click “Add Property…”
4. For key, enter “hive.metastore.schema.verification” and for value enter “false”
5. Select “Add” once you are finished and then restart the appropriate services.

**Start Hue**

1. Remember, you must start Hue before you can access Hue in your browser.
2. cd [space]
3. Once you are in your home directory, you will run [hue.sh] to start Hue

**Creating and Querying a Table in Hive**

There are a number of ways to interact with Hive. In this exercise you will use the Hive Query Editor in Hue.

1. Visit the Hue page in Firefox.

2. Open Hive query editor, by selecting the editor of your choice from the Query Editors menu.

1. In the query editor pane on the right, enter an SQL command to create a table for the web page data imported in previous exercises:

CREATE EXTERNAL TABLE webpage (page\_id SMALLINT,

name STRING, assoc\_files STRING)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ‘\t'

LOCATION ‘/user/training/lanier/webpage’

1. Click the **Execute** button to execute the command.
2. To see the table you just created, refresh the table list on the left.
   * Note: In the Hue Query Editor you will be prompted with the type of refresh to perform; choose **Clear Cache** and click **Refresh.**
3. Click on the **webpage** table to see the column definitions. (You may need to scroll down through the list of tables to see the columns.)
4. Click the **New Query** button, then enter and execute a test query such as:

SELECT \* FROM webpage WHERE name LIKE "ifruit%"

The resulting data is shown in the **Results** tab of the pane below the query.

1. Click on the **Preview Sample Data** icon. (Note: Hover your mouse pointer over the table to view the controls.)

**[OPTIONAL] Using Sqoop to Import Directly into Hive**

In this section you will use Sqoop to import data from MySQL into HDFS and

also automatically create the corresponding table in the Hive Metastore.

9. In a terminal window, import the device table directly into the Hive Metastore.

$sqoop import

--connect jdbc:mysql://localhost/lanier

--username [username] —password [password]

--fields-terminated-by '\t'

--table device

**--hive-import**

Use --hive-import for either Hive; this adds metadata to the Metastore, which both tools use.

NOTE:You may get a warning message thatchgrpis unable to change theownership of the generated files; you can disregard the warning, it does not affect the import.

1. Using Hue or the command line, review the imported data files. The Sqoop Hive import copies the data to the default Hive warehouse location:

/user/hive/warehouse/device.

1. If you are using Impala, refresh the Impala metadata cache by entering the command in the Hue Query Editor:

INVALIDATE METADATA

12. As in the previous exercise, view the columns and execute a test query:

SELECT \* FROM device LIMIT 10

**END**